



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/662,502	09/15/2003	Mu Li	M61.12-0527	9194

27366 7590 03/27/2007  
WESTMAN CHAMPLIN (MICROSOFT CORPORATION)  
SUITE 1400  
900 SECOND AVENUE SOUTH  
MINNEAPOLIS, MN 55402-3319

EXAMINER

SERROU, ABDELALI

ART UNIT	PAPER NUMBER
----------	--------------

2626

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
3 MONTHS	03/27/2007	PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

**Office Action Summary**

Application No.

10/662,502

Applicant(s)

LI ET AL.

Examiner

Abdelali Serrou

Art Unit

2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-31 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-31 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 15 September 2003 is/are: a) ☐ accepted or b) ☒ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |   |   |
|---|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. ____. |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                  | 5) <input type="checkbox"/> Notice of Informal Patent Application                       |
| 3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08)<br>Paper No(s)/Mail Date ____. | 6) <input type="checkbox"/> Other: ____.  |

## **DETAILED ACTION**

### ***Information Disclosure Statement***

1. The information disclosure statement filed on 5/27/04 fails to comply with 37 CFR 1.98(a)(3) because it does not include a concise explanation of the relevance, as it is presently understood by the individual designated in 37 CFR 1.56(c) most knowledgeable about the content of the information, of each patent listed that is not in the English language. It has been placed in the application file, but the information referred to therein has not been considered.

### ***Drawings***

2. The drawings are objected to as failing to comply with 37 CFR 1.84(p)(5) because they include the following reference character(s) not mentioned in the description: Element "502" in figure 5 (should be "508"). Corrected drawing sheets in compliance with 37 CFR 1.121(d), or amendment to the specification to add the reference character(s) in the description in compliance with 37 CFR 1.121(b) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the changes are not accepted by the examiner, the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

*Claim Objections*

3. **Claim 1** is objected to because of the following informalities:

Claim 1 preempts “ a computer readable medium including instructions readable by a computer”, which does not reflect the intended scope of the claim. To further timely prosecution, the Examiner interpreted this phrase as “ a computer readable medium storing instructions readable by a computer”.

Appropriate correction is required.

*Claim Rejections - 35 USC § 101*

4. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

**Independent claim1, and 14** and by dependency claims 2-13, and 15-24, are rejected under 35 U.S.C. 101 because independent claims 1 and 14 do not fall within one of the four categories patentable subject matter of 35 U.S.C § 101 (process, machine, manufacture, or composition of matter).

**Independent claim 1** preempts a computer readable medium. According to the specification (pages 7 – 8) the claimed computer readable media comprises communication media that embodies a modulated data signal such as a carrier wave. A carrier wave signal is nothing but the physical characteristics of a form of energy, such as a frequency, voltage, or the strength of a magnetic field. A signal, interpreted as an abstract Idea, is a subject matter that is not a practical application or use of an idea, a law of nature or a natural

phenomenon, and so is not patentable. See, e.g., *Rubber-Tip Pencil Co. v. Howard*, 87 U.S. (20 Wall.) 498, 507 (1874) (“idea of itself is not patentable, but a new device by which it may be made practically useful is”); *Mackay Radio & Telegraph Co. v. Radio Corp. of America*, 306 U.S. 86, 94, 40 USPQ 199, 202 (1939) (“While a scientific truth, or the mathematical expression of it, is not patentable invention, a novel and useful structure created with the aid of knowledge of scientific truth may be.”); *Warmerdam*, 33 F.3d at 1360, 31 USPQ2d at 1759 (“steps of 'locating' a medial axis, and 'creating' a bubble hierarchy . . . describe nothing more than the manipulation of basic mathematical constructs, the paradigmatic 'abstract idea'”). See *Le Roy v. Tatham*, 55 U.S. (14 How.) 156, 175. (1852) (“A principle, in the abstract, is a fundamental truth; an original cause; a motive; these cannot be patented, as no one can claim in either of them an exclusive right.”); *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127, 132, 76 USPQ 280, 282 (1948) (combination of six species of bacteria held to be nonstatutory subject matter).

**Independent claim 14** preempts an abstract idea, as evidenced by computer readable medium claim 1. Computer programs claimed as computer listings per se, i.e., the descriptions or expressions of the programs are not physical “things.” They are neither computer components nor statutory processes, as they are not “acts” being performed. Such claimed computer programs do not define any structural and functional interrelationships between the computer program and other claimed elements of a computer, which permit the computer program’s functionality to be realized. In contrast, a claimed computer-readable medium encoded with a computer program is a computer element which defines structural and functional interrelationships between the computer program and the rest of the computer which permit the

Art Unit: 2626

computer program's functionality to be realized, and is thus statutory. See Lowry, 32 F.3d at 1583-84, 32 USPQ2d at 1035.

Furthermore, **claim 14** does not claim a practical application of the claimed language model, i.e., being used in a translation, speech recognition, or speech synthesis system. If the "acts" of a claimed process manipulate only numbers, abstract concepts or ideas, or signals representing any of the foregoing, the acts are not being applied to appropriate subject matter. Benson, 409 U.S. at 71-72, 175 USPQ at 676. Thus, a process consisting solely of mathematical operations, i.e., converting one set of numbers into another set of numbers, does not manipulate appropriate subject matter and thus cannot constitute a statutory process.

(See: Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility).

Accordingly, the subject matter of claims 1-14 is held to be nonstatutory subject matter.

### ***Claim Rejections - 35 USC § 103***

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

**Claims 1-4, 6-7, 14, 15-21, 23, 25-26, and 28** are rejected under 35 U.S.C. 103(a) as being unpatentable over Chen et al. (U.S 5,806,021 issued on Sept. 8, 1998) (hereinafter: Chen)

Art Unit: 2626

in view of Brockett et al. (U.S. 6,968,308, filed Nov. 1, 2000 and issued on Nov. 22, 2005) (hereinafter: Brockett).

**As per claims 1, 14, and 25**, Chen teaches segmenting the sentence into two possible segmentations (col. 3, lines 18-32);

performing a Forward Maximum Matching (FMM) segmentation (col. 3, lines 37-65) and a Backward Maximum Matching (BMM) segmentation (col. 3, line 66 - col. 4, line 24);

generating an n-gram model (col. 4, lines 45-47), and

selecting one of the two segmentations as a function of probability information for the two segmentations (col. 4, lines 25-26).

Chen does not explicitly teach recognizing an overlapping ambiguity string in the input sentence as a function of the two segmentations, and replacing the overlapping ambiguity string with tokens.

Brockett in the same field of endeavor teaches recognizing the overlapping ambiguity string in the input sentence as a function of the two segmentations (col. 2, lines 16-17), and replacing the overlapping ambiguity string with tokens (inherent in selecting the most segmentation for the input string (col. 11, lines 5-19).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine the overlapping ambiguity string recognizer of Brockett to the text segmentation system of Chen, because Brockett suggests that this would better identify the right segment among the competing segments (col. 1, lines 55-63).

**As per claims 2-4, 23, and 26**, Chen in view of Brockett teach obtaining the probability information from a lexical knowledge base (lexicon, col. 2, line 41), wherein the lexical

Art Unit: 2626

knowledge base comprises a trigram model (col. 45-49), wherein selecting one of the two segmentations comprises classifying the probability information (col. 3, lines 29-32, wherein the probability information (likelihood) of both segmentations is calculated and classified to select the segmentation with higher likelihood).

**As per claims 6-7, and 28**, Chen teaches performing a Forward Maximum Matching (FMM) segmentation, for recognizing a segmentation  $O_f$ , (col. 3, lines 37-65) and a Backward Maximum Matching (BMM) segmentation for recognizing a segmentation  $O_b$  of the input sentence (col. 3, line 66 - col. 4, line 24).

Chen does not explicitly teach recognizing an overlapping ambiguity string in the input sentence as a function of the two segmentations.

Brockett in the same field of endeavor teaches recognizing the overlapping ambiguity string in the input sentence as a function of the two segmentations (col. 2, lines 16-17).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine the overlapping ambiguity string recognizer of Brockett to the text segmentation system of Chen, because Brockett suggests that this would better identify the right segment among the competing segments (col. 1, lines 55-63).

**As per claim 13**, Chen teaches wherein the unsegmented language is Chinese (col. 3, line 21).

**As per claim 15**, Chen teaches determining a probability associated with each of the FMM segmentation of the overlapping ambiguity string and the BMM segmentation of the overlapping ambiguity string (col. 3, lines 18-32).



As per claims 16-18, Chen teaches an N-gram model (col. 4, lines 45-47), and probability information about a first and last word of the overlapping ambiguity string (col. 5, lines 1-5, wherein probability of each part of the phrase (word), resulted from a segmentation is compared separately).

As per claims 19-21, Chen teaches N-gram model (col. 4, lines 45-47), that uses information about a string of words comprising a first word of the overlapping ambiguity string and two context words to the left of the first word, and a last word of the overlapping ambiguity string and two context words to the right of the last word (inherently disclosed in the process of determining likelihood scores using n-grams models (tri-gram model), col. 5, lines 45-47).

6. Claims 5, 8-12, 22, 24, 27, and 29-31, are rejected under 35 U.S.C. 103(a) as being unpatentable over Chen in view of Brockett, as applied to claims 4, 15, and 23, and further in view of Pedersen ("*A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation*", in Proceedings of the First Annual Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 63-69, April 29 – May 4, 2000).

As per claim 5, 22, and 24, Chen in view of Brockett teaches all the limitations of claims 4, 15, and 23, upon which claims 5, 22, and 24 depend.

Chen and Brockett do not explicitly teach using an ensemble of Naive Bayesian Classifiers.

Pederson in the same field of endeavor teaches using an ensemble of Naive Bayesian Classifiers (Abstract).

Art Unit: 2626

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine Pederson's Nave Bayesian Classifier with the automatic text segmenter of Chen, because Pederson suggests that this would provide more accurate disambiguation systems (Abstract).

**As per claims 8-12**, Chen in view of Brockett teach one of the two segmentations (col. 4, lines 25-26), classifying the probability information of  $O_f$  and  $O_b$  (col. 3, lines 29-32, wherein the probability information (likelihood) of both segmentations is calculated and classified to select the segmentation with higher likelihood), and determining which one of the said probabilities is higher (col. 4, lines 25-26).

Chen and Brockett do not explicitly selecting one of the two segmentations is a function of a set of context features, words around the overlapping ambiguity string, associated with the overlapping ambiguity string, classifying the probability information of the context features, and determining which one of the said probabilities is higher, as a function of the set of context features.

Pederson in the same field of endeavor teaches the Naïve Bayesian Classifier for word sense disambiguation based on windows of context (Pages 63-64).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to use the Naïve Bayesian Classifier of Pederson in combination with the text segmenting system of Chen, to use the probability information of the context features to select one of the two segmentations. Pederson suggests that this would provide more accurate disambiguation systems (Abstract).

**As per claims 27 and 29**, Chen in view of Brockett teaches all the limitations of claims 25 and 28, upon which claims 27 and 29 depend.

Chen and Brockett do not explicitly teach generating an ensemble of classifiers as a function of an n-gram model.

Pederson in the same field of endeavor teaches generating an ensemble of classifiers as a function of an n-gram model (Abstract, and page 64, col. 2, lines 15-19).

Therefore, it would have been obvious to a person of ordinary skill in the art at the time of the invention was made to combine Pederson's classifiers with the combined system of Chen and Brockett, because Pederson suggests that this would provide more accurate disambiguation systems (Abstract).

**As per claim 30**, Chen, Brockett, and Pederson teach all the limitations of claim 29, upon which claim 30 depends. Chen in view of Brockett, furthermore, teach approximating probabilities of the FMM and BMM segmentations of each overlapping ambiguity string as being equal to the product of individual unigram probabilities of individual words in the FMM and BMM segmentations respectively, of the overlapping ambiguity string (col. 3, line 37 –col. 4, line 26, wherein the probabilities of the FMM and BMM segmentations of each overlapping ambiguity are approximated and compare to choose the one with the highest score).

**As per claim 31**, Chen, Brockett, and Pederson teach all the limitations of claim 30, upon which claim 31 depends. Pederson, furthermore, teach a joint probability of a set of context features conditioned on an existence of one of the segmentations of each overlapping ambiguity string (ambiguous word) as a function of a corresponding probability of a leftmost and a

Art Unit: 2626

rightmost word of the corresponding overlapping ambiguity string (Pages 63-64, 2<sup>nd</sup> paragraph, NaiveBayesian Classifiers).

### *Conclusion*

7. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Bai et al. (U.S 6,311,152) teach a system for Chinese tokenization and named entity recognition. Gao et al. (U.S 2004/0243408) teach a method and apparatus using source-channel models for word segmentation. Mackie (U.S 2003/0097252) teaches a method and apparatus for efficient segmentation of compound words using probabilistic breakpoint traversal. Kaji et al. (U.S 4,750,122) teach a method for segmenting a text into words. Chu (U.S 6,374,210) teaches an automatic segmentation of a text. Lin (U.S 6,620,207) teaches a method and apparatus for processing Chinese teletext. Wu et al. (U.S 6,640,006) teach word segmentation in chinese text. Wu et al. (U.S 6,678,409) teach parameterized word segmentation of unsegmented text. Zamora (U.S 5,448,474) teaches a method for isolation of Chinese words from connected Chinese text. Luo et al. (Proceedings of the 19th international conference on Computational linguistics, 2002, Volume 1, pp: 1-7) teach covering ambiguity resolution in Chinese word segmentation based on contextual information.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Abdelali Serrou whose telephone number is 571-272-7638. The examiner can normally be reached on 8:30-5:00.

Art Unit: 2626

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis I. Smits can be reached on 571-272-7628. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

A. Serrou  
3/22/07



**DAVID HUDSPETH**  
SUPERVISORY PATENT EXAMINER  
TECHNOLOGY CENTER 2600